A proof-of-concept study: exon-level expression profiling and alternative splicing in autism using lymphoblastoid cell lines

Zohreh Talebizadeh[a], Richard Aldenderfer[a] and Xue Wen Chen[b]

Objective **Autism is a complex, heterogeneous neurobehavioral disorder with many causes and varying degrees of severity. Some genetic implications related to autism may involve gene-regulatory processes such as alternative splicing. Here, we assess the feasibility of profiling exon-level gene expression in autism using the Affymetrix Human exon 1.0 ST array.**

Methods **We examined lymphoblastoid cell line-derived RNAs from five patients with autism compared with five controls.**

Results **Analysis of variance and Bonferroni multiple test correction identified 57 genes exhibiting differential exon-level expression, suggesting potential changes in the resultant alternatively spliced transcripts in autism compared with controls. Genes with differentially expressed exons included *CYFIP1*, a previously reported autism susceptibility gene. Furthermore, several genes recently reported to have deregulated alternative splicing in autism brain samples showed differential exon expression in our autism group.**

Conclusion **The paucity of autism brain samples and extensive phenotypic heterogeneity of autism demands finding ways to also identify autism-related genomic events in accessible nonbrain resources, which may contribute in biomarker identifications. This proof-of-concept study shows that the analysis of alternative splicing in lymphoblastoid cell line samples has a potential to reveal at least a subset of brain-related deregulation of splicing machinery that might be implicated in autism. *Psychiatr Genet* 24:1–9 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

Psychiatric Genetics 2014, **24**:1–9

Keywords: alternative splicing, autism, differential expression, exon array, lymphoblastoid cell lines

[a]Section of Medical Genetics and Molecular Medicine, Children's Mercy Hospital and University of Missouri–Kansas City School of Medicine, Kansas City, Missouri and [b]Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, Kansas, USA

Correspondence to Zohreh Talebizadeh, PhD, Children's Mercy Hospitals and Clinics, 2401 Gillham Road, Kansas City, MO 64108, USA
Tel: +1 816 983 6506; fax: +1 816 983 6501;
e-mail: ztalebi@cmh.edu

Received 3 July 2012 **Revised** 26 December 2012 **Accepted** 7 March 2013

Introduction

Autism is a genetically heterogeneous early-onset common neurodevelopmental disorder. It encompasses a group of conditions known as autism spectrum disorders (ASD), which includes classical autism, pervasive developmental disorder–not otherwise specified, and Asperger syndrome. Diagnostic features for classical autism include significant impairment in three developmental domains: reciprocal social behavior, language development, and repetitive/stereotypic behaviors (Lord *et al.*, 1994). Despite significant progress in the field of autism research, including the identification of rare mutations, the underlying genetic factors remain to be identified in the majority of cases.

Alternative splicing of primary transcripts (pre-mRNAs) is a regulatory mechanism contributing to the increased complexity of higher eukaryotic organisms resulting in structurally and functionally distinct transcript isoforms and protein variants. Up to 74% of human multiexonic genes undergo alternative splicing (Johnson *et al.*, 2003). These transcript isoforms are produced through a combination of processes such as the selection of cassette exons, mutually exclusive exons, intron retention, exon skipping, alternative 5′ or 3′ splice sites, and alternative promoters or polyadenylation sites (Blencowe, 2006).

The exon array is a high-throughput method used to study alternative splicing by evaluating global gene expression at the exon level (Lee and Roy, 2004). Affymetrix exon arrays have been used to examine alternative splicing and differential exon expression in diseases such as cancer and to study the transition of human embryonic stem cells to neural stem cells. Yeo *et al.* (2007) used exon array to compare three types of human stem cells and observed alternative splicing changes during neuronal differentiation. Similarly, evaluation of 13 human mid-fetal brain regions showed differential exon usage suggestive of brain region-dependent splicing (Johnson *et al.*, 2009). Some of the genes identified as differentially expressed in brain regions (i.e. *CNTNAP2*, *ROBO1*) are associated with neurological disorders such as autism and dyslexia (Anitha *et al.*, 2008; Bakkaloglu *et al.*, 2008; Poelmans *et al.*, 2009). More recently, an exon array analysis was used to show exon expression and alternative splicing in blood samples from individuals with Tourette syndrome compared with healthy controls (Tian *et al.*, 2011).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (*www.psychgenetics.com*).

0955-8829 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/YPG.0b013e3283635526

Despite the growing evidence for the regulatory influence of alternative splicing on gene expression, particularly in brain function, this process has not been evaluated extensively in ASD. More recently, transcriptomic analysis using RNA-seq showed splicing deregulation in autistic brains compared with controls (Voineagu *et al.*, 2011). Although an essential piece in identifying the underlying molecular mechanisms of autism is to examine brain samples, the other remaining piece of this complex condition is how to deal with the existing phenotypic heterogeneity. Undoubtedly, a key challenge in identifying biomarkers and therapeutic interventions is this extensive heterogeneity in the presentation of ASD. A potential approach to address this issue would be to focus on a homogenous group of autistic patients and to detect brain-related biomarkers from accessible resources, including whole blood or lymphoblastoid cell lines (LCL) that can be used to investigate the genetic underpinnings of ASD subsets.

Fresh blood is a preferred choice compared with Epstein–Barr virus (EBV)-transformed cell lines for genetic studies. However, currently, LCLs are the main available biomaterials from major autism repositories such as the Autism Genetic Resource Exchange (AGRE) to be used by the investigators. Therefore, the key question is to know to what extent LCL samples can be informative in genetic studies. To address this critical gap in the area of alternative splicing, we evaluated global exon expression in LCL-derived RNAs from autistic patients compared with unaffected controls. Our objective was to determine whether evaluation of LCL samples has the potential to detect differentially expressed alternatively spliced isoforms in autism susceptibility genes. Our rationale was that evaluation of the expression level of exons from LCL, a highly accessible nonbrain tissue, allows high-throughput analysis of at least a subset of brain-related alternative splicings that are also deregulated in LCL. Such a method can serve as a model to investigate specific patterns of misplicing in relation to the etiology of autism.

Materials and methods

Participants

EBV-transformed LCLs from peripheral lymphocytes were obtained from the AGRE (Los Angeles, California, USA). The autistic group included three females (6, 11, and 13 years old) and two males (5 and 12 years old) diagnosed with classical autism (one from a simplex and four from multiplex families). The control group included five age-matched and sex-matched unaffected AGRE siblings unrelated to the selected autistic patients. Additional AGRE samples (15 autistic females; 12 control females) were used for exon array verification assays. Brain autopsy samples for control participants, available from the NIH Brain and Tissue Bank, from four regions (frontal cortex, cerebrum, cerebellum, and brainstem), were used to confirm the presence of detected novel alternatively spliced isoforms in brain tissue. ZT is an approved AGRE investigator. Samples used in this study were deidentified and thus considered 'exempt' from IRB review.

Total RNA preparation and exon array profiling

Total RNA was isolated from each cell line using TRIzol Reagent with Phase Lock Gel-Heavy (Invitrogen, Carlsbad, California, USA) according to the manufacturer's instruction. The same RNA was utilized for exon arrays and the follow-up experiments.

Array preparation and profiling was performed by the University of Kansas Medical Center microarray core facility. In brief, RNA was prepared using the GeneChip Whole Transcript Sense Target Labeling Assay (Affymetrix, Santa Clara, California, USA). For each sample, ribosomal RNA (rRNA) reduction was performed using RiboMinus (Invitrogen) on 1 μg of total RNA. Following rRNA reduction, double-stranded cDNA was synthesized with random hexamers tagged with a T7 promoter sequence. The tagged cDNA was used as a template for amplification with T7 RNA polymerase to generate cRNA. Next, random hexamers were used to reverse transcribe the cRNA to produce single-stranded DNA. The single-stranded DNA was fragmented and then labeled with terminal deoxynucleotidyl transferase. A hybridization cocktail was prepared and hybridized to Affymetrix GeneChip Human Exon 1.0 ST arrays. Processed arrays were scanned using the GeneChip Scanner 3000 7G with Autoloader (Affymetrix) to assess whole-genome differential gene/exon expression and alternative splicing.

Microarray data analysis

The Affymetrix Human exon 1.0 ST array contains ~5.4 million probes grouped into 1.4 million probe sets, interrogating over 1 million exon clusters (i.e. about four probes per exon with a median of 30–40 probes per transcript) (Affymetrix, 2006). In Affymetrix exon array data processing, multiple probes targeting a particular region (e.g. a given exon) are grouped to form a probe set. Probe sets are assembled into transcript clusters and correspond to every possible isoform transcribed from a single gene. For simplicity, we refer to transcript clusters as genes in the rest of this paper.

Affymetrix Expression Console Software (Affymetrix) was used to derive exon-level expression values from the CEL file probe-level hybridization intensities using the model-based Robust Multi-chip Analysis (RMA) algorithm (Irizarry *et al.*, 2003). RMA performs normalization (by adjusting the baseline intensities to the median of the control samples), background correction, and data summarization of probe set intensities. The RMA-normalized data were subsequently analyzed to visualize differential exon expression using GeneSpring GX 11.0 (Agilent Technologies, Foster City, California, USA). The data analyzed were then filtered to only include probe sets

in the 'core' meta-probe list, representing 17 800 RefSeq genes and full-length GenBank mRNAs.

All probe sets defined as 'present' were filtered on the basis of a *P*-value cutoff of 0.05 generated during the execution of the detected above-background (DABG) algorithm. On the basis of DABG *P*-values, genes with at least 60% of core exons being present in at least 60% of samples (three of five) in at least one of two conditions (autism or control group) were retained. Next, analysis of variance (ANOVA) was carried out on individual probe sets that are present in at least 60% of samples in at least one of two conditions. ANOVA *P*-values were multitest corrected using the Bonferroni method (i.e. dividing unadjusted *P*-values by the number of multiple tests, using a cutoff value of adjusted $P < 0.05$).

Exon array data were also analyzed with tools available in Genomics Suite software (v6.6; Partek Inc., St Louis, Missouri, USA) using Affymetrix annotation files [NetAffx, version na31. hg19 (Affymetrix)]. The RMA algorithm was used for probe set (exon level) intensity analysis for background correction, quintile normalization with baseline transformation to median of the control samples. Adjustments were made for GC content and probe sequence on prebackground-subtracted values. Preanalysis filtration included the following two steps: (a) removing any probe set that is not expressed in at least one sample group using DABG *P*-value less than 0.05 as cutoff and (b) including probe sets that had a group mean of $\log_2$ signal greater than 3. Alternative splice detection was performed using a two-way ANOVA. The analysis is performed at the exon level, but the result is displayed at the transcript level. All genes represented by less than 3 probe sets in the transcripts were removed, as it is often difficult to interpret alternative exon incorporation patterns with so few markers. Any transcripts not represented by a HUGO gene symbol were also removed, maintaining the focus of the analysis on known genes.

Verification of exon array data

Genes showing differentially expressed exons identified by exon arrays were evaluated for select biologically relevant genes using quantitative assays [quantitative reverse transcription PCR (qRT-PCR) and TaqMan], followed by DNA sequencing. All qRT-PCR reactions were run in triplicate on an ABI 7000 using SuperScript III Platinum SYBR Green One-Step qRT-PCR w/ROX (Invitrogen/Applied Biosystems, Carlsbad, California, USA). The comparative C_t method (Applied Biosystems, 1997) with *GAPDH* as the endogenous control was used to calculate gene expression for both SYBR Green and TaqMan assays. *GAPDH* may not be the best internal control for all tissues, and *GUSB* has been suggested as a better choice of a reference gene for studying LCL-derived samples (De Brouwer *et al.*, 2006). Running *GAPDH* and *GUSB* on our samples did not show any noticeable differences in the normalized data. Therefore, qRT-PCR data are shown using the commonly used internal control (i.e. *GAPDH*).

All transcripts' sequences were verified with direct DNA sequencing of amplicons using the BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies/Applied Biosystems) on an ABI 3100 Avant Genetic Analyzer (Life Technologies/Applied Biosystems). Primer and probe information can be found in Supplementary Table 1, Supplemental Digital Content 1 (*http://links.lww.com/PG/A78*).

Pathway analyses

Pathway analysis was carried out using the Database for Annotation, Visualization, and Integrated Discovery version 6.7 (DAVID, *http://david.abcc.ncifcrf.gov/*), a publicly available functional annotation tool (Dennis *et al.*, 2003; Irizarry *et al.*, 2003; Huang da *et al.*, 2007).

Results

Exon array data analysis

Exon arrays contain sequences from two main sources: high-confidence mRNAs (core set) and a lower confidence set of theoretical genes. For this study, we carried out data analysis for the core set transcripts using GeneSpring. In doing so, from 17 800 core transcripts, 10 072 transcripts (57%) fulfilled the applied filtration criteria and were identified as being expressed above background levels. Within this gene set, ANOVA and Bonferroni multiple test correction were used to identify exon-level differential expression. The use of this highly stringent method for the 10 072 genes showed 57 genes to be differentially expressed at the exon level between autism and control samples (Table 1). Information on corrected and uncorrected *P*-values and fold changes is listed in Supplementary Tables 4 and 5, Supplemental Digital Content 2 and 3 (*http://links.lww.com/PG/A79* and *http://links.lww.com/PG/A80*). Age and sex are the two potential confounding factors in our study. The study groups examined were age-matched and the limited sample size does not allow evaluations of sex-specific differences. Therefore, no potential confounding factors were observed.

Partek software was used to further validate differential exon expression in these 57 genes. Expression patterns of probe sets were inspected manually to find exons expressed differentially with respect to the rest of the gene. After this manual inspection of probe set expressions, three genes (*CYFIP1*, *TRAP150*, and *ZMYM6*) were selected for RT-PCR validations on the basis of the aforementioned criteria and biological relevance. Graphic presentations for these three genes generated by Partek are shown in Supplementary Fig. 1, Supplemental Digital Content 6 (*http://links.lww.com/PG/A83*). The first two genes showed differential expression of a probe set in the middle of a gene, which is considered to be a more reliable indicator of spliced isoforms than the ones toward

Table 1 Identified genes ($n=57$) showing differential exon-level expression in autism compared with control samples (with *P*-value < 0.05 after Bonferroni correction)

Gene symbol	Chromosome	Gene symbol	Chromosome	Gene symbol	Chromosome
CSDE1	1p13[a,b]	KIAA0746	4p15[b]	DDEF1	8q24[a,b]
AGL	1p21[b]	HTT	4p16[a,b]	SMARCA2	9p24[b]
SLC30A7	1p21[b]	ALPK1	4q25	ITGB1	10p11[b]
MTF2	1p22[b]	NIPBL	5p13[a,b]	ZMIZ1	10q22[b]
NSUN4	1p33[b]	RNASEN	5p13[a,b]	C12orf57	12p13[b]
MACF1	1p34[b]	TARS	5p13[a,b]	LRP6	12p13[b]
TRAP150	1p34[b]	PPWD1	5q12[b]	SLC38A1	12q13[a,b]
ZMYM6	1p34[b]	MRPS27	5q13	KNTC1	12q24[b]
KIF1B	1p36[b]	LMNB1	5q23[b]	MED13L	12q24[b]
BPNT1	1q41	C6orf89	6p21[b]	KPNA3	13q14[b]
RAB3GAP2	1q41	TRIM27	6p22[b]	DACT1	14q23[b]
MTR	1q43[b]	ASCC3	6q16[a,b]	CYFIP1	15q11[a,b]
EFR3B	2p23	REV3L	6q21[b]	TTLL13	15q26[b]
PPM1G	2p23	RNF217	6q22[b]	PLCG2	16q23[b]
SLC35A5	3q13[b]	EZR	6q25[b]	RICH2	17p12[b]
RPN1	3q21[b]	VPS41	7p14[b]	AKAP1	17q22[a,b]
CHST2	3q23[a]	AHR	7p21[b]	WDR68	17q23[a,b]
SLC9A9	3q24[b]	HOOK3	8p11[b]	SEPT9	17q25[a,b]
NDUFB5	3q26[a,b]	PRKDC	8q11[b]	HCK	20q11

P-values (corrected and uncorrected), probe sets IDs, and fold changes are listed in Supplementary Tables 4 and 5 (*http://links.lww.com/PG/A79* and *http://links.lww.com/PG/A80*).
[a]Suggestive linked region has been reported for this cytoband [see citations in Supplementary Table 3 (*http://links.lww.com/PG/A87*)]; 13/57 = 23%.
[b]Autism-related copy number variation has been reported for this cytoband [source: Autism Genetic Database (Matuszek and Talebizadeh, 2009)]; 49/57 = 86%.

the 5′ or 3′ end. For the third gene, a probe set distance from the middle was intentionally selected to also evaluate the accuracy of differential expression for this type of probe sets.

Differential expression of *CYFIP1* and detection of new alternatively spliced variants

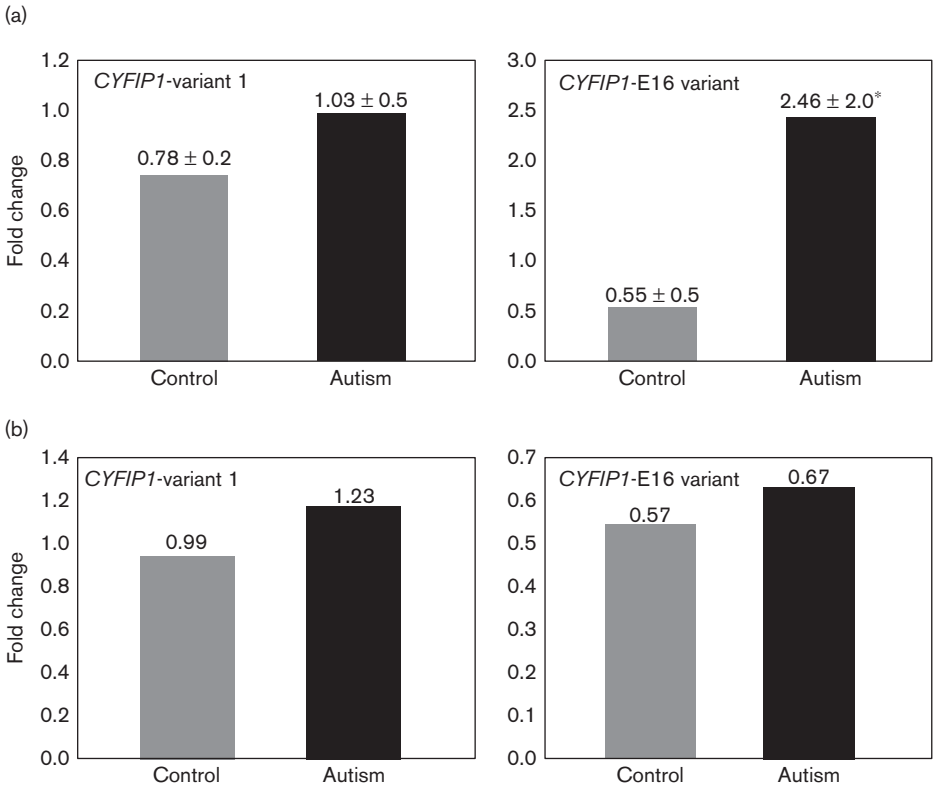
There are two protein-coding transcripts in the RefSeq database for *CYFIP1*: variant 1 (NM_014608) encoding a long form and variant 2 (NM_001033028) encoding a short form. Exon array analysis of *CYFIP1* showed a higher expression for probe set 3583680, which binds to exon 16 in variant 1 (corresponding to exon 2 in variant 2), in autistic patients versus controls. We designed two sets of RT-PCR primers to detect variants 1 and 2 separately using the sequence divergence upstream of this exon. RT-PCR, followed by DNA sequencing for variant 1 also detected a product missing exon 16 (E16 isoform) introducing a premature stop codon. qRT-PCR showed a higher expression of variant 1 in autism compared with control samples, but no changes were observed for variant 2. TaqMan assays confirmed higher expression in autistic patients (using the same RNA samples utilized in exon array) for both *CYFIP1* variants 1 and E16 (Fig. 1a). There was a wider variation in expression values (i.e. fold change calculated on the basis of a single reference sample) in the autism group (i.e. 0.5, 1.3, 1.1, 2.5, and 5.9 in three females and two males, respectively) compared with the controls. The overall expression was significantly higher in the autistic patients compared with the control samples ($P = 0.02$).

Although not statistically significant, a trend toward higher expression of both variants 1 and E16 was also detected on evaluating additional LCL samples using pooled cDNA (15 autistic females vs. 12 unaffected females). The TaqMan assay results are shown in Fig. 1b. The use of pooled samples masks expression variations within each study group. However, for this proof-of-concept study, a pooled strategy was used to assess the presence of a trend in line with the exon array data, not to detect individual variations. Furthermore, because of the existing heterogeneity in this population, it is less likely to replicate a finding without patient stratifications and a large sample size. Most likely, this is why the significantly higher expression detected in the original five autistic patients (Fig. 1a) was not observed in the pooled samples (Fig. 1b).

A series of RT-PCR reactions, followed by DNA sequencing were run for *TRAP150* (NM_005119) and *ZMYM6* (NM_007167) to determine whether differentially expressed exons detected in the array are subject to alternative splicing. DNA sequencing of the amplified products confirmed that these exons undergo alternative splicing and resulted in the identification of previously unreported alternative splicing isoforms for these two genes [i.e. no record in either Ensembl (Hubbard *et al.*, 2009) or Hollywood (Holste *et al.*, 2006)]. These new variants included one isoform for *TRAP150* (missing exon 4 resulting in an in-frame loss of 301 amino acid residues) and five alternatively spliced *ZMYM6* variants that are labeled on the basis of the missing exons (i.e. isoforms missing exon 2; exon 4; exons 2 and 4; exons 2 and 5; and exons 2 and 4 and 5). *ZMYM6* variants missing exon 2, the location of the start codon, most likely do not code for proteins and the exclusion of only exon 4 introduces a premature stop codon.

Electropherogram images and a schematic illustration of the evaluated probe sets and detected splice isoforms are shown in Supplementary Fig. 2, Supplemental Digital Content 7 (*http://links.lww.com/PG/A84*). Expressions of the

Fig. 1



Expression of *CYFIP1* using TaqMan assay. TaqMan assays were used to compare the expression of the full-length *CYFIP1* transcript (variant 1) and a novel isoform missing exon 16, between autism and control samples. The samples used in these quantitative experiments were (a) original LCL samples used in our exon array (five autism and five controls) and (b) additional LCL samples (pooled cDNA from 15 autistic and 12 control females). To calculate gene expression, all samples were calibrated to a given control sample and normalized to *GAPDH*. Only the E16 variant in the original samples [marked with an asterisk in (a)] showed differential expression in autism compared with controls ($P=0.02$). The remaining comparisons were not significantly different between the two study groups. LCL, lymphoblastoid cell line.

novel alternatively spliced isoforms for these three genes (*CYFIP1*, *TRAP150*, and *ZMYM6*) were confirmed in brain samples using RT-PCR or TaqMan assays. The amplification curves for brain samples are shown in Supplementary Fig. 3, Supplemental Digital Content 8 (*http://links.lww.com/PG/A85*).

Gene pathway analysis

The 57 genes showing Bonferroni-corrected differential exon expression (Table 1) were uploaded into DAVID to assess gene function and relevant pathways. Only one biological pathway (protein-lipid modification; PANTHER-BP00070) passed the threshold for multiple testing corrections (Bonferroni-corrected $P = 0.03$). Six out of 57 (11%) genes including *CYFIP1* and *ZMYM6* were included in this pathway. *SMARCA2*, a schizophrenia-associated gene (Christoforou *et al.*, 2007), was also among the six genes enriched in this biological pathway. The remaining three genes in this pathway from Table 1 were *PPWD1*, *MTF2*, and *KIAA0746*.

Discussion

A wide range of neurological disorders have been associated with the deregulation of RNA processing including genetic forms of Parkinson's disease (Citron *et al.*, 2002), epilepsy (Heinzen *et al.*, 2007), Alzheimer's disease (Citron *et al.*, 2002; Heinzen *et al.*, 2007; Dahmcke *et al.*, 2008), Rett syndrome (Abuhatzira *et al.*, 2005), and Prader–Willi syndrome (Kishore and Stamm, 2006). Within the nervous system, alternative splicing allows cells to continually adjust their protein balance to respond and adapt to different stimuli (Licatalosi and Darnell, 2006). It also appears to play a dominant role in regulating neuronal gene expression and function (Black and Grabowski, 2003). Another example of the role that alternative splicing plays in brain function can be observed in neuregulin genes. Neuregulins are involved in crucial aspects of neural biology and are expressed in the mammalian nervous system. It has been shown that particular splicing forms of neuregulin genes show different expression patterns in the human embryonic central nervous system than in the adult brain (Bernstein *et al.*, 2006; Carteron *et al.*, 2006).

A few studies have shown the importance of alternative splicing in autism. *PRKCB1*-spliced isoforms were reported to be significantly downregulated in the autistic brain (Lintas *et al.*, 2009). Structural and splice variants of neurexins, presynaptic membrane cell-adhesion molecules, have been suggested as a predisposing factor for autism (Yan *et al.*, 2008). Previously, we reported novel alternative splicing isoforms in two X-linked neuroligins, *NLGN3* and *NLGN4*, and discussed the potential role of alternative splicing in the etiology of autism (Talebizadeh *et al.*, 2006a). More recently, dysregulated splicing in the ASD brain was observed using RNA sequencing (Voineagu *et al.*, 2011).

In the Ensembl human database, the number of genes with more than one annotated transcript is smaller than the estimated number of alternatively spliced genes. This discrepancy suggests that there are likely many undetected or unreported alternative splicing events not represented in the database. Therefore, in contrast to more traditional expression arrays that target the 3′ end of a transcript, exon arrays can identify novel transcriptional events (Okoniewski *et al.*, 2007).

Here, we assessed the feasibility of using LCL samples to study global exon expression as an indicator for alternative splicing profiling in autism. Among the differentially expressed exons identified in our study are genes previously associated with autism (*CYFIP1*) or related to nervous system development, including *TRAP150* and *ZMYM6*. Further evaluation of these three genes resulted in the identification of new alternatively spliced isoforms. The novel splice isoforms detected in this study were also expressed in brain, indicating that they are not LCL-specific transcripts. As these splicing isoforms may result in different encoded proteins or a truncated transcript, their expression level may potentially be important to the etiology of diseases.

CYFIP1 maps to the chromosome 15q11 Prader–Willi syndrome region encoding the cytoplasmic FMR1 interacting protein 1 (Schenck *et al.*, 2003). Upregulation of *CYFIP1* has been reported in autism patients with maternal duplications of 15q11 using genome-wide expression profiling of LCLs (Nishimura *et al.*, 2007), suggesting a potential molecular link between *FMR1* and *CYFIP1*. Cosegregation of a chromosome 15q11 microduplication with autism in a Dutch pedigree and overexpression of the *CYFIP1* gene, residing in the duplicated region, in the autistic patients further pinpoints this gene as an autism susceptibility gene (Van der Zwaag *et al.*, 2010). Jiang *et al.* (2008) reported that *CYFIP1* undergoes complex alternative splicing including exon skipping and the usage of two different promoters. Our pilot study shows that evaluation of alternatively spliced *CYFIP1* isoforms is a more sensitive method than measurement of total mRNA to detect potential changes in this autism susceptibility gene in autistic patients.

Furthermore, such approaches may unveil new autism susceptibility genes not identified by conventional gene expression assays. For example, the current pilot alternative splicing study provides evidence that *TRAP150* and *ZMYM6* might be considered as potentially new autism candidate genes and should be further assessed in follow-up studies. *TRAP150* containing the mammalian transcriptional regulatory complex TRAP (thyroid hormone receptor-associated protein) is a component of the spliceosome and can trigger mRNA degradation in the nucleus (Zhou *et al.*, 2002; Lee *et al.*, 2010). *TRAP220*, another member of the TRAP gene family, has been associated with *RELN* and *BDNF*, two autism-related genes (Sui *et al.*, 2010). *ZMYM6* has tandem MYM (myeloproliferative and mental retardation) zinc figure domains (Smedley *et al.*, 1999). Its gene homologue (*ZMYM3*), expressed most abundantly in the brain (Scheer *et al.*, 2000), has been disrupted in a patient with X-linked mental retardation (Van der Maarel *et al.*, 1996).

To assess that differentially expressed exons identified in our study are not specific to LCLs, similar data analysis was used to exon array data using LCLs from an independent set of participants (five patients with a chromosomal disorder and five controls). There was no overlap between the Bonferroni-corrected gene lists generated from these two data sets, suggesting that array data in our pilot study are unlikely to be an artifact because of the LCL-derived RNA and may potentially have some relevance to the etiology of autism.

Evaluation of differences in gene expression between males and females using exon array methods on CEPH LCLs did not identify sex-specific splicing variants (Zhang *et al.*, 2009). The limited sample size in our study does not allow evaluation of sex-specific differences; however, comparison of expression data in females showed several genes including *CDKL5* to be differentially expressed in our three autistic females compared with control females. *CDKL5* is an X-linked gene and mutations in this gene have been reported in Rett syndrome, a condition that occurs almost exclusively in females (Weaving *et al.*, 2004). Given the small sample size, interpretation of this result warrants some caution, but it is worth noting that genome-wide linkage studies and clinical evaluation of autistic patients have provided evidence for differences at both the genotypic and the phenotypic level between males and females. It remains to be seen whether there is also sex-specific alternative splicing profiling in autism.

Differential alternative splicing of 212 genes has been identified recently in autism brain samples using high-throughput RNA sequencing (RNA-Seq) (Voineagu *et al.*, 2011). None of the 57 Bonferroni-corrected genes overlap with the reported autism brain transcriptome profile. However, when uncorrected differentially expressed exons from the main DABG list were examined, we

found differential exon expression for 78 of these genes (i.e. 37%) in our autism group compared with the controls [Supplementary Table 2a, Supplemental Digital Content 4 (*http://links.lww.com/PG/A81*)]. The direction of the changes was not checked between the two data sets (i.e. RNA-seq on brain samples and our LCL-derived exon array data). Ten of 78 genes with only one or two differentially expressed exons were further assessed for the correlation of the corresponding exons in the two studies. In five of these 10 genes, we observed that the same exon was differentially expressed in both LCL and autism brain samples [Supplementary Fig. 4, Supplemental Digital Content 9 (*http://links.lww.com/PG/A86*)]. Given the small sample size in our pilot study, finding this level of overlap with brain profiles for differentially expressed transcripts is promising and further suggests that a subset of brain-expressed transcripts may be detected in LCL.

To further check for the potential correlation of gene expression between LCL and the brain, we examined gene expression of previously reported up/downregulated genes in autism brain samples. Eleven genes reported previously by Garbett *et al.* (2008) to be differentially expressed in autism brain showed a similar expression direction in the LCL autism samples examined in our study [Supplementary Table 2b, Supplemental Digital Content 4 (*http://links.lww.com/PG/A81*)].

Although examination of total mRNA expression has been valuable in identifying upregulated or downregulated genes in autism, it does not provide information about potential changes in mRNA alternative splicing levels and the resulting isoforms. It has been shown that gene expression and alternative splicing profiling are independent and may not be correlated for a given set of genes (Yeo *et al.*, 2004; Clark *et al.*, 2007), a significant finding further emphasizing the importance and necessity of evaluating alternative splicing patterns in autism.

In the present study, comparison of exon array profiles between a small group of autism and control participants identified several genes with potential differential expression at the exon level in LCL samples. In agreement with our previous work on neuroligins (Talebizadeh *et al.*, 2006a) and recent transcriptomic analysis of autistic brain (Voineagu *et al.*, 2011), the current study provides further evidence that changes at the alternative splicing level should be considered in the etiology of autism.

A systematic comparison of exon arrays and RNA-seq profiles from identical participants showed a strong concordance between these genomic methods at both gene-level and exon-level expression, suggesting that microarrays remain useful and accurate for transcriptomic analysis, particularly for detection of expression changes of low-abundance genes (Raghavachari *et al.*, 2012). Comparison of the result in the current proof-of-concept exon array study and reported brain RNA-seq gene expressions (Voineagu *et al.*, 2011) suggested that, in the absence of having access to a large number of autism brain samples, evaluation of LCL samples may be informative in detecting changes for a subset of brain-expressed alternative splicing isoforms.

Limitations of this pilot study

The present pilot study has a number of methodological limitations, including the use of LCL samples, heterogeneity and sex differences in ASD, and replication issue.

(a) LCL samples may not be the ideal system to investigate brain-related genes; however, they may still be informative for at least a subset of brain-related changes. Rollins *et al.* (2010) compared whole-genome expression in blood, LCL, and brain samples to assess whether nonbrain biomaterials can be informative in the field of biomarker discovery for neuropsychiatric disorders. Array results showed about *22%* coexpression levels of transcripts between postmortem patients' brain and blood profiles. A similar correlation was detected on comparing exon arrays from rat brain and blood samples (19%). Global transcriptome analysis of blood and EBV-transformed cell lines was also carried out in the same patients before and after transformation to determine the effects of EBV transformation on the gene expression profiles. More than 60% of transcripts expressed in equivalent levels in blood and LCL samples. A majority of the differentially expressed genes between blood and LCL were located at six specific cytobands (1p, 1q, 2p, 2q, 7q, and 19q), indicating the presence of regional hotspots of gene dysregulation following EBV infection. Correspondence of some of these sites to reported EBV integration sites further supports the nonrandom integration of EBV into the genome (Gao *et al.*, 2006). The low interparticipant variability observed among healthy individuals, comparing LCL and blood gene expression profiles, indicates a uniform impact by EBV through transformation (Gao *et al.*, 2006; Rollins *et al.*, 2010).

To find a list of potentially associated genes with EBV, we searched Gene Cards (*http://genecards.org/index.shtml*) using the keyword EBV, which returned 471 genes. Only one of the 57 genes listed in Table 1, *EZR*, was found to be a potentially EBV-associated gene on the basis of the Gene Cards search. Also, none of the previously reported EBV-related hotspots by Rollins *et al.* (2010) (either cytobands or genes) were observed among our Bonferroni-corrected genes. Taken together, these assessments suggest that in the present pilot study, detected differential expressions in LCL samples from autism samples are less likely to be because of EBV infections.

(b) We are aware of the limitation of using pooled samples because it does not allow detection of individual variations within each group. Given the known

heterogeneity in autism and the observed variation in the individual expression values in the original five autistic patients, differential expression of *CYFIP1* isoforms may only exist in a subset of autistic participants, not all. Detection of such a susceptible autistic subset requires investigation of the potential role of isoforms in larger study cohorts coupled with subtyping approaches. This is why the detected difference in the original five samples (Fig. 1a) is not replicated in the 15 additional samples using a pooling approach (Fig. 1b). Furthermore, normal distribution of gene isoform expression levels should be established in control populations to enable detection of disease-related profiles. Gene expression data from 98 autopsy samples (11 different tissues) of control participants (48 females and 50 males), examined in a previously published human expression profile (Son *et al.*, 2005; Talebizadeh *et al.*, 2006b), showed a low degree of variation in the overall expression of *CYFIP1* [Supplementary Table 6, Supplemental Digital Content 5 (*http://links.lww.com/PG/A82*)]. The female to male ratio (Talebizadeh *et al.*, 2006b) did not show a drastic contrast (i.e. ranging between 0.8 and 1.3) in the overall expression of this gene in each tissue (Son *et al.*, 2005). It should be noted that such a total gene expression does not allow detection of expression profiles (or sex bias) of different *CYFIP1* isoforms, which needs to be addressed in future studies.

(c) With the exception of a few highly confirmed and replicated genes, the rest of the autism genetic findings need to be validated. Heterogeneity is one of the main obstacles contributing toward the challenges in replicating such findings. However, this issue is beyond the scope of our current proof-of-principle study as the objective here is to show the feasibility of using exon array in autism studies, not to identify disease-causing genes, which requires large sample sizes.

(d) It should also be noted that there might be a limitation in the analysis with respect to the adjustment for multiple testing. The Bonferroni correction assumes that the multiple tests are independent; thus, it is too stringent for gene expression studies that are aiming to find biologically relevant factors or connected networks. As the objective of the current study is a proof of principle, not detecting autism-related expression profile, we used Bonferroni corrections to observe whether any biologically relevant findings can be found in the small list with a low rate of false-negative results.

Acknowledgements

The authors wish to thank Dr Nataliya Kibiryeva (CMH) for assisting with exon array data analysis. The authors acknowledge the AGRE families; without them, this research would not have been possible. They also gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to Clara M. Lajonchere (PI). Partial funding support was provided from the Children's Mercy Hospital-KBR Award (CMH/01.4232) and Autism Speaks (2578) to Z.T.

The authors thank the Kansas University Medical Center-Microarray Facility (KUMC-MF) for generating array data sets. The Microarray Facility is supported by the Kansas University-School of Medicine, KUMC Biotechnology Support Facility, the Smith Intellectual and Developmental Disabilities Research Center (HD02528), and the Kansas IDeA Network of Biomedical Research Excellence (RR016475). Human tissues were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland, Baltimore, Maryland under contracts NO1-HD-4-3368 and NO1-HD-4-3383.

Conflicts of interest

There are no conflicts of interest.

References

Abuhatzira L, Makedonski K, Galil YP, Gak E, Ben Zeev B, Razin A, Shemer R (2005). Splicing mutation associated with Rett syndrome and an experimental approach for genetic diagnosis. *Hum Genet* **118**:91–98.

Affymetrix (2006). Identifying and validating alternative splicing events. Available at: *http://media.affymetrix.com/support/technical/technotes/id_altsplicingevents_technote.pdf*.

Anitha A, Nakamura K, Yamada K, Suda S, Thanseem I, Tsujii M, *et al.* (2008). Genetic analyses of roundabout (ROBO) axon guidance receptors in autism. *Am J Med Genet B Neuropsychiatr Genet* **147B**:1019–1027.

Applied Biosystems (1997). User Bulletin #2: ABI PRISM 7700 Sequence Detection System. Available at: *http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_040980.pdf*.

Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, *et al.* (2008). Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet* **82**:165–173.

Bernstein HG, Lendeckel U, Bertram I, Bukowska A, Kanakis D, Dobrowolny H, *et al.* (2006). Localization of neuregulin-1alpha (heregulin-alpha) and one of its receptors, ErbB-4 tyrosine kinase, in developing and adult human brain. *Brain Res Bull* **69**:546–559.

Black DL, Grabowski PJ (2003). Alternative pre-mRNA splicing and neuronal function. *Prog Mol Subcell Biol* **31**:187–216.

Blencowe BJ (2006). Alternative splicing: new insights from global analyses. *Cell* **126**:37–47.

Carteron C, Ferrer-Montiel A, Cabedo H (2006). Characterization of a neural-specific splicing form of the human neuregulin 3 gene involved in oligodendrocyte survival. *J Cell Sci* **119**:898–909.

Christoforou A, Le Hellard S, Thomson PA, Morris SW, Tenesa A, Pickard BS, *et al.* (2007). Association analysis of the chromosome 4p15-p16 candidate region for bipolar disorder and schizophrenia. *Mol Psychiatry* **12**:1011–1025.

Citron BA, Suo Z, SantaCruz K, Davies PJ, Qin F, Festoff BW (2002). Protein crosslinking, tissue transglutaminase, alternative splicing and neurodegeneration. *Neurochem Int* **40**:69–78.

Clark TA, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H, *et al.* (2007). Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol* **8**:R64.1–R64.16.

Dahmcke CM, Buchmann-Moller S, Jensen NA, Mitchelmore C (2008). Altered splicing in exon 8 of the DNA replication factor CIZ1 affects subnuclear distribution and is associated with Alzheimer's disease. *Mol Cell Neurosci* **38**:589–594.

De Brouwer AP, van Bokhoven H, Kremer H (2006). Comparison of 12 reference genes for normalization of gene expression levels in Epstein–Barr virus-transformed lymphoblastoid cell lines and fibroblasts. *Mol Diagn Ther* **10**:197–204.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4;** pp. 3–11.

Gao J, Luo X, Tang K, Li X, Li G (2006). Epstein–Barr virus integrates frequently into chromosome 4q, 2q, 1q and 7q of Burkitt's lymphoma cell line (Raji). *J Virol Methods* **136**:193–199.

Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K, Persico AM (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol Dis* **30**:303–311.

Heinzen EL, Yoon W, Weale ME, Sen A, Wood NW, Burke JR, *et al.* (2007). Alternative ion channel splicing in mesial temporal lobe epilepsy and Alzheimer's disease. *Genome Biol* **8**:R32.1–R32.18.

Holste D, Huo G, Tung V, Burge CB (2006). HOLLYWOOD: a comparative relational database of alternative splicing. *Nucleic Acids Res* **34**: D56–D62.

Huang da W, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, *et al.* (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* **35**:W169–W175.

Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, *et al.* (2009). Ensembl 2009. *Nucleic Acids Res* **37**:D690–D697.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**:249–264.

Jiang YH, Wauki K, Liu Q, Bressler J, Pan Y, Kashork CD, *et al.* (2008). Genomic analysis of the chromosome 15q11-q13 Prader–Willi syndrome region and characterization of transcripts for GOLGA8E and WHCD1L1 from the proximal breakpoint region. *BMC Genomics* **9**:50.

Johnson JM, Castle J, Garrett-Engele P, Kan Z, Loerch PM, Armour CD, *et al.* (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**:2141–2144.

Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D, *et al.* (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**:494–509.

Kishore S, Stamm S (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**:230–232.

Lee C, Roy M (2004). Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol* **5**:231.1–231.4.

Lee KM, Hsu Ia W, Tarn WY (2010). TRAP150 activates pre-mRNA splicing and promotes nuclear mRNA degradation. *Nucleic Acids Res* **38**:3340–3350.

Licatalosi DD, Darnell RB (2006). Splicing regulation in neurologic disease. *Neuron* **52**:93–101.

Lintas C, Sacco R, Garbett K, Mirnics K, Militerni R, Bravaccio C, *et al.* (2009). Involvement of the PRKCB1 gene in autistic disorder: significant genetic association and reduced neocortical gene expression. *Mol Psychiatry* **14**:705–718.

Lord C, Rutter M, Le Couteur A (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24**: 659–685.

Matuszek G, Talebizadeh Z (2009). Autism Genetic Database (AGD): a comprehensive database including autism susceptibility gene-CNVs integrated with known noncoding RNAs and fragile sites. *BMC Med Genet* **10**:102.

Nishimura Y, Martin CL, Vazquez-Lopez A, Spence SJ, Alvarez-Retuerto AI, Sigman M, *et al.* (2007). Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum Mol Genet* **16**:1682–1698.

Okoniewski MJ, Yates T, Dibben S, Miller CJ (2007). An annotation infrastructure for the analysis and interpretation of Affymetrix exon array data. *Genome Biol* **8**:R79.1–R79.9.

Poelmans G, Engelen JJ, Van Lent-Albrechts J, Smeets HJ, Schoenmakers E, Franke B, *et al.* (2009). Identification of novel dyslexia candidate genes through the analysis of a chromosomal deletion. *Am J Med Genet B Neuropsychiatr Genet* **150B**:140–147.

Raghavachari N, Barb J, Yang Y, Liu P, Woodhouse K, Levy D, *et al.* (2012). A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. *BMC Med Genomics* **5**:28.

Rollins B, Martin MV, Morgan L, Vawter MP (2010). Analysis of whole genome biomarker expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet* **153B**:919–936.

Scheer MP, van der Maarel S, Kubart S, Schulz A, Wirth J, Schweiger S, *et al.* (2000). DXS6673E encodes a predominantly nuclear protein, and its mouse ortholog DXHXS6673E is alternatively spliced in a developmental- and tissue-specific manner. *Genomics* **63**:123–132.

Schenck A, Bardoni B, Langmann C, Harden N, Mandel JL, Giangrande A (2003). CYFIP/Sra-1 controls neuronal connectivity in Drosophila and links the Rac1 GTPase pathway to the fragile X protein. *Neuron* **38**:887–898.

Smedley D, Hamoudi R, Lu YJ, Cooper C, Shipley J (1999). Cloning and mapping of members of the MYM family. *Genomics* **60**:244–247.

Son CG, Bilke S, Davis S, Greer BT, Wei JS, Whiteford CC, *et al.* (2005). Database of mRNA gene expression profiles of multiple human organs. *Genome Res* **15**:443–450.

Sui L, Ren WW, Li BM (2010). Administration of thyroid hormone increases reelin and brain-derived neurotrophic factor expression in rat hippocampus in vivo. *Brain Res* **1313**:9–24.

Talebizadeh Z, Lam DY, Theodoro MF, Bittel DC, Lushington GH, Butler MG (2006a). Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. *J Med Genet* **43**:e21.1–e21.7.

Talebizadeh Z, Simon SD, Butler MG (2006b). X chromosome gene expression in human tissues: male and female comparisons. *Genomics* **88**:675–681.

Tian Y, Liao IH, Zhan X, Gunther JR, Ander BP, Liu D, *et al.* (2011). Exon expression and alternatively spliced genes in Tourette Syndrome. *Am J Med Genet B Neuropsychiatr Genet* **156B**:72–78.

Van der Maarel SM, Scholten IH, Huber I, Philippe C, Suijkerbuijk RF, Gilgenkrantz S, *et al.* (1996). Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. *Hum Mol Genet* **5**:887–897.

Van der Zwaag B, Staal WG, Hochstenbach R, Poot M, Spierenburg HA, de Jonge MV, *et al.* (2010). A co-segregating microduplication of chromosome 15q11.2 pinpoints two risk genes for autism spectrum disorder. *Am J Med Genet B Neuropsychiatr Genet* **153B**:960–966.

Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, *et al.* (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**:380–384.

Weaving LS, Christodoulou J, Williamson SL, Friend KL, McKenzie OL, Archer H, *et al.* (2004). Mutations of CDKL5 cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *Am J Hum Genet* **75**:1079–1093.

Yan J, Noltner K, Feng J, Li W, Schroer R, Skinner C, *et al.* (2008). Neurexin 1alpha structural variants associated with autism. *Neurosci Lett* **438**:368–370.

Yeo G, Holste D, Kreiman G, Burge CB (2004). Variation in alternative splicing across human tissues. *Genome Biol* **5**:R74.1–R74.15.

Yeo GW, Xu X, Liang TY, Muotri AR, Carson CT, Coufal NG, Gage FH (2007). Alternative splicing events identified in human embryonic stem cells and neural progenitors. *PLoS Comput Biol* **3**:1951–1967.

Zhang W, Duan S, Bleibel WK, Wisel SA, Huang RS, Wu X, *et al.* (2009). Identification of common genetic variants that account for transcript isoform variation between human populations. *Hum Genet* **125**:81–93.

Zhou R, Bonneaud N, Yuan CX, de Santa Barbara P, Boizet B, Schomber T, *et al.* (2002). SOX9 interacts with a component of the human thyroid hormone receptor-associated protein complex. *Nucleic Acids Res* **30**:3245–3252.